HIGH SCHOOL EDITION
Journal of Student Research

Volume 13 Issue 1 (2024)

CardioRetinaNet: Joint Analysis of Retinal Images for Cardiovascular Disease Diagnosis and Age Estimation Using Convolutional Neural Network

Donghun Won[1] and Mary Cruz[#]

[1]Skyline High School, USA
[#]Advisor

ABSTRACT

In recent years, despite the consistent progress in the field of medicine, a rapid growth of patients with cardiovascular disease has been a problem globally. Cardiovascular disease, a general term for disorders of the heart or blood vessels, require high-speed and precise detection to control and avoid hostile consequences. Traditionally, computerized tomography scans, utilizing images retrieved from CT-scans, to calculate coronary artery calcium score were the dominant method for the diagnosis of cardiovascular diseases. This usual technique is problematic because of its lengthy procedure and the difficulty of getting the diagnosis due to the expensive cost. To address this problem, I propose a convolutional neural network-based cardiovascular disease diagnosis system using retinal images which are notably more cost-effective than computerized tomography scans. The proposed system takes retinal images as input and generates a categorical assessment of the severity of cardiovascular disease as output. Additionally, it provides a predicted age of the patient, contributing to an enhanced performance in cardiovascular disease classification. By incorporating these features, the proposed system aims to advance the accuracy of cardiovascular disease diagnosis. The experimental results clearly demonstrate that the proposed system attains a state-of-the-art performance in diagnosing cardiovascular disease. I expect that the proposed system will make a significant contribution to the utilization of retinal images as a biomarker for diagnosis of cardiovascular disease.

Introduction

Cardiovascular disease (CVD) is the disease of heart and blood vessels and it remains as one of the leading causes of mortality and morbidity worldwide. The early detection and accurate diagnosis of CVD is important in improving patient outcomes and reducing the burden on healthcare systems. Improvement of CVD diagnosis is a heavier concept nowadays because of the soaring number of CVD patients in the U.S. With the escalating patients, efficiency is the primary objective.

Traditionally, the diagnosis of CVD has been conducted by measuring coronary artery calcium score (CACS) from CT scan images. The quantity of CACS is the main risk determinant and medically verified indicator of CVD. But this accustomed method has a handful of issues such as having a low accuracy (giving vague scores for patients with the disease when they should get a higher score), delayed diagnosis, and its price (CT scans are expensive).

As stated before, many downsides are shown in the traditional method of CVD diagnosis. To address this problem, numerous studies have been conducted on AI-powered biomarkers for the diagnosis of CVD. These methods frequently leverage CNN architectures to analyze medical images, such as retinal images or CT scans, as CNN can effectively learn intricate patterns and features from large datasets, enabling accurate and efficient image interpretation. De Vos et al. introduces a computationally efficient approach for automated calcium scoring in CT scans, utilizing two convolutional neural networks (De Vos et al. 2019). The first network aligns input CTs' fields of view, while the second network directly predicts the calcium score, yielding accurate results in real time and offering potential application in clinical and research settings. Rim et al. developed a deep learning-based cardiovascular risk stratification system using retinal photographs to predict CACS (Rim et al. 2021). The system's accuracy in predicting cardiovascular events surpasses traditional methods.

Inspired by these, I proposed a novel convolutional neural network-based cardiovascular disease diagnosis system using retinal images. The proposed system utilizes retinal images as input and delivers a categorical evaluation of disease severity as its output. Moreover, the system outputs a predicted age of the patient, augmenting its capability to achieve superior performance in cardiovascular disease classification.

The structure of this paper is as follows: Chapter 2 presents background knowledge, Chapter 3 details the proposed method, Chapter 4 outlines the experimental results, and Chapter 5 concludes the study.

Related Work

Cardiovascular Disease

Cardiovascular Disease (CVD) refers to disorder of the heart and blood vessels. With CVD being the global leading cause of death, its detection rate is the key to the lives of millions of people. Because heart disease is critical to one’s body after it has become active, early detection is especially important to prevent further damage.

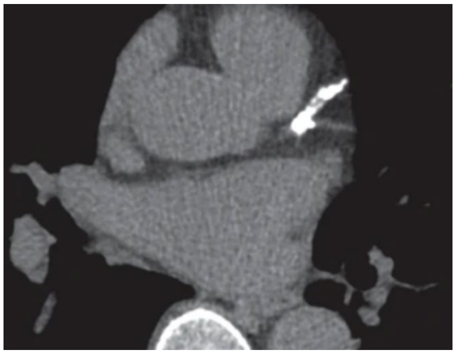

Figure 1. Example of calcified coronary plaques in CT scan

In the medical field, CVD has been discovered through CT-scans by calculating the CACS. But with the high price and lengthy time, only a small portion of patients were able to get their treatment at the right time. Recently, the use of artificial intelligence technology on the diagnosis of CVD has been receiving the spotlight because of its quick diagnosis time and fair price. This new technique is able to achieve this by using other relatively inexpensive medical images or data such as retinal or x-ray images.

Image Classification

Image classification is a core computer vision task where it receives an image as an input and returns the image assigned to one of a fixed set of categories. Image classification is implemented with the usage of neural network or conventional neural network (CNN) systems. These networks utilize the loss function to calculate the errors, gradient descent algorithm to train the model, and activation functions such as max function to create multiple layers. Probability of each input is calculated with the cross-entropy loss function with the softmax function.

Some of the previous models that have used the architecture of CNN are AlexNet (Krizhevsky et al. 2012), VGG (Simonyan et al. 2014), and Resnet (He et al. 2016) . For example, assume there is a CNN model that classifies different kinds of animals. If the final probability for a cat image is 0.998, the model has an outstanding quality of precision.

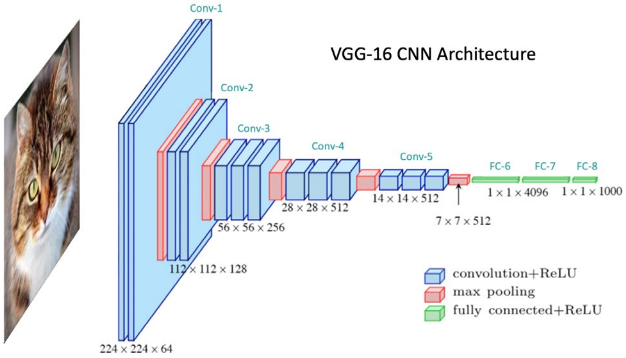

Figure 2. Example VGG-16 (Simonyan et al. 2014) convolutional neural network architecture (LearnOpenCV 2023)

Proposed Method

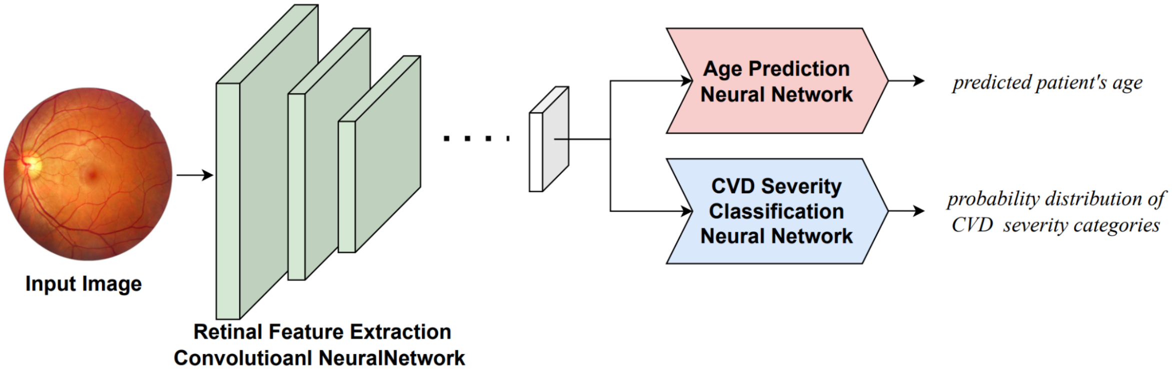

Figure 3. Architecture of the proposed cardiovascular diagnosis system

Figure 3 shows the overall architecture of the proposed model. The input image is a retinal image which is observed by the Retinal Feature Extraction Convolutional Neural Network. The input retinal image $\boldsymbol{I}$ is inputted to the proposed Retinal Feature Extraction where the feature map z is produced. Here I define this process as follows: $RFE: I \rightarrow z$. The random variable I denotes the retinal image that is to be assessed and z denotes the output feature map produced from the input image. Then, feature map z is fed to both Age Prediction Neural Network and CVD Severity Classification Neural Network. The predicted patient's age and the probability distribution of CVD severity categories are each returned as the results, respectively. These would be the key factors in CVD diagnosis. The proposed model uses joint analysis by utilizing the regression method and the classification method. The final prediction of age is accomplished with the regression method.

To train the proposed method, I utilize cross-entropy loss function and mean squared error function. Cross-entropy loss measures the performance of a classification model whose output is a probability value between 0 and 1. Cross-entropy loss increases as the predicted probability diverges from the actual label. It is commonly used in classification problems. Mean squared error function is often used in regression problems. In regression, the goal is to predict a continuous quantity (patients' age). Mean squared error function measures the average squared difference between the predicted values and the actual values.

Equation 1: Mean Squared Error Function

$$L_{MSE}=\frac{1}{N}\sum_{i=1}^{N}\left|Y_i-P_i\right|^2$$

Here Y_i, and P_i denote the ground truth, and predicted value of the model, respectively. The variable N represents the total number of samples in the dataset. To find the mean squared error value, the squared difference of the observed value (ground truth) and the predicted value - known as the ‘error’ of the model - is calculated and this is iterated for all observations. Then the sum of all the squared values of the errors divided by the total number of samples.

Equation 2: Cross-Entropy Loss Function

$$L_{CE} = -\ln(P(y))$$

In equation 2, $P(y)$ denotes the predicted probability of severity category with regard to the ground truth category. Finally, the total loss function is defined as equation 3.

Equation 3: Total Loss Function

$$L_{\blacksquare} = L_{MSE} + L_{CE}$$

For the training parameters, I set the number of epochs to 100. An epoch represents one complete pass through the entire training dataset. The batch size is set to 512. This parameter determines the number of samples processed before the model's parameters are updated. The initial learning rate is set to 0.0001. The learning rate controls the size of the steps taken during the optimization process. To further optimize the learning process, I employ a learning rate schedule. The learning rate is decreased by a factor of 0.1 at epochs 40 and 80.

Experimental Results

CACS Dataset

Table 1. Sample distribution of CACS dataset

	CACS	Number of patients	Percentage
Absent	0	13,243	53.8%
Mild	1~100	5,871	23.8%
Moderate	101~400	3,434	14.0%
Severe	>400	2,079	8.4%
Total	-	24,627	100%

Table 1. shows the number of patients out of 24,627 patients for each category divided by certain ranges of CACS. 53.8 percent of the total patients fill up the category of 0 CACS, which means they do not have any risk of CVD yet. if the score is from 0 to 100, the patient has a low risk of CVD and is still relatively normal. The

average age of the patients in the dataset is 59.2 where 16171 (65.7%) patients are male and 8456 (34.3%) patients are female.

Experimental Protocol

By using the K-fold cross validation method, more accurate results can be expected from the dataset. This method splits the total dataset into *k* folds to run *k* experiment cases. In each distinct case, one of the folds is chosen for validation, meaning it will be the only group for testing while the others will be only used for training the model. If the average accuracy of each of the experiment cases are similar, it is considered to be high in quality because of the low variance of the performance evaluation. By training with this method to measure the quality of the model on new data, more data can be used efficiently to train the model and it also helps solve the overfitting problem.

Evaluation with State-of-the-Art Methods

Table 2. Evaluation with state-of-the-art methods

	Accuracy	Recall	Precision	F1-Score
VGG19 (Simonyan et al. 2014)	0.8556 (±0.0010)	0.8025 (±0.008)	0.8402 (±0.0012)	0.8145 (±0.0014)
MobileNetV2 (Sandler et al. 2018)	0.8578 (±0.0014)	0.8043 (±0.0009)	0.8389 (±0.0008)	0.8169 (±0.0010)
EfficientNet-B7 (Tan et al. 2019)	0.8608 (±0.0008)	0.8078 (±0.0007)	0.8520 (±0.0014)	0.8208 (±0.0009)
HRNet-40 (Wang et al. 2020)	0.8720 (±0.0016)	0.8190 (±0.0014)	0.8524 (±0.0008)	0.8321 (±0.0012)
Resnet-50 (He et al. 2016)	0.8762 (±0.0010)	0.8236 (±0.0012)	0.8565 (±0.0010)	0.8361 (±0.0009)
Proposed Method (Resnet-50 based)	0.8948 (±0.008)	0.8418 (±0.0012)	0.8754 (±0.0013)	0.8548 (±0.0012)

Table 2 presents the error and the accuracy of the proposed model as well as the other state-of-the-art methods. All of the models have their recall value as the lowest, meaning that the false positive compared to the true positive was relatively high when contrasted with the other values. As the state-of-the-art models progress, the results show constant improvement. The accuracy of the VGG19 model was 0.8556 ± 0.0010, which improved to the value of 0.8948 ± 0.008 of our Resnet-50 based proposed model. Progression of 3.92% for accuracy, 3.93% for recall, 3.52% for precision, and 4.03% for the F1 score were shown. As the goal is to improve the accuracy of the model, the improvements of 0.24%, 0.39%, 1.13%, 0.40%, and 1.87% in the F1 score results from 81.45% to 85.48% showed consistency.

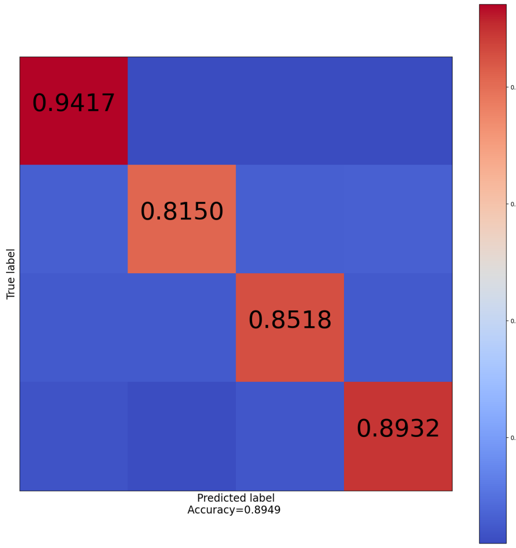

Figure 4. Confusion Matrix

Figure 4 above displays the confusion matrix of the proposed model with the four categories of the classification based on the CACS levels. The degree of the percentages are represented by the colors red and blue, the percentage written showing when the model's predicted label paralleled the actual label. Confusion matrix is a summary of the performance of the model represented in a matrix form. With the red-colored squares in the matrix that show the true positives and false negatives, the degree of performance is summarized easier.

Ablation Study

Table 3. Ablation study (without age-guided approach)

	Accuracy (ablation)	**Accuracy (full model)**
VGG19 (Simonyan et al. 2014)	0.8556 (±0.0010)	0.8688 (±0.0014)
MobileNetV2 (Sandler et al. 2018)	0.8578 (±0.0014)	0.8716 (±0.0013)
EfficientNet-B7 (Tan et al. 2019)	0.8608 (±0.0008)	0.8770 (±0.0009)
HRNet-40 (Wang et al. 2020)	0.8720 (±0.0016)	0.8877 (±0.0012)
Resnet-50 (He et al. 2016)	0.8762 (±0.0010)	0.8948 (±0.008)

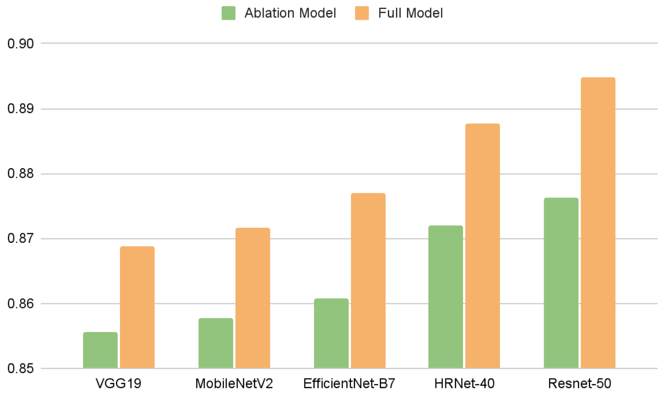

Figure 5. Ablation study results (without age-guided approach)

To increase the accuracy and the level of performance of the model, the proposed model is applied on various kinds of other CNN models. These models with the application of the proposed method are compared to observe the results with higher accuracy. In Table 3 and Figure 5, The full model based on Resnet-50 created the highest results out of the state-of-the-art models.

Table 4. Data augmentation experiment results

	Accuracy
Baseline	0.8948 (±0.008)
Grayscale	0.8847 (±0.007)
Sharpness	0.8482 (±0.0014)
Color Jitter	0.8587 (±0.0009)
CHALE	0.8856 (±0.0017)
Rotation (+)	0.9103 (±0.0010)
Horizontal Flip (+)	0.9005 (±0.0011)

In addition to the ablation study, data augmentation study is also performed to enhance the model's accuracy further. Augmentation methods are important since they are used to reduce the problem of overfitting and improve the model's performance. By modifying the dataset of images, data augmentation technique gives us a more diversified dataset. When applied to the original dataset, it would advance the model's quality by making it able to classify numerous different kinds of images.

In this study, after the ablation study and several various techniques, 7 kinds of data augmentation techniques are utilized. These include: augmentations with the tools of grayscale, sharpness, color jitter, CHALE, rotation, and horizontal flip.

Conclusion

In this research, I have presented a convolutional neural network system for cardiovascular disease (CVD) diagnosis based on the application of previously introduced convolutional neural network models, such as HRNet-40 and Resnet-50. While the traditional way was using the CT-scans as inputs, the introduced model utilizes retinal images. By using retinal images, the diagnosis is cost-effective, needs less manpower, fast, and it is non-invasive. The model is structured of the feature extractor convolutional neural network that gets the retinal images as input, then the Age Prediction Neural Network and CVD Severity Classification Neural Network receive the resulting layer of the convolutional neural network. To increase the accuracy of the model and minimize the errors, k-fold cross validation technique, mean squared error function, and cross entropy loss functions were used. The proposed model was applied to 5 kinds of state-of-the-art methods. From the ablation study, the CNN network of Resnet-50 showed the highest performance. For more various kinds of data, 7 data augmentation techniques were used.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

De Vos, B. D., Wolterink, J. M., Leiner, T., De Jong, P. A., Lessmann, N., & Išgum, I. (2019). Direct automatic
coronary calcium scoring in cardiac and chest CT. IEEE transactions on medical imaging, 38(9), 2127-2138.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). https://doi.org/10.48550/arXiv.1512.03385

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25.

LearnOpenCV. (2023, Jan 18). *"Understanding Convolutional Neural Network (CNN): A Complete Guide"*: LearnOpenCV https://learnopencv.com/understanding-convolutional-neural-networks-cnn/

Rim, T. H., Lee, C. J., Tham, Y. C., Cheung, N., Yu, M., Lee, G., ... & Wong, T. Y. (2021). Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. The Lancet Digital Health, 3(5), e306-e316.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520). https://doi.org/10.48550/arXiv.1801.04381

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning (pp. 6105-6114). PMLR. https://doi.org/10.48550/arXiv.1905.11946

Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., ... & Xiao, B. (2020). Deep high-resolution representation learning for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3349-3364. https://doi.org/10.48550/arXiv.1908.07919